R E P O R T R E S U M E S

ED 015 654 48 EM 004 043

ANSWER OBSERVING IN PROGRAMMED INSTRUCTION--II. THE EFFECT OF INCENTIVE ON STUDENT PERFORMANCE WITHIN AND AFTER PROGRAMMED INSTRUCTION.

BY- GEIS, GEORGE L. NIELSEN, SUSAN
MICHIGAN UNIV., ANN ARBOR

REPORT NUMBER BR-6-1784-3
CONTRACT OEC-3-6-061784-0503
EDRS PRICE MF-$0.25 HC-$0.96 22P.

DESCRIPTORS- *PROGRAMED INSTRUCTION, *LEARNING MOTIVATION, *REINFORCERS, PROGRAMING PROBLEMS

THIS STUDY REPLICATES EARLIER DATA CONCERNING CONDITIONS UNDER WHICH STUDENTS CHECK OR "OBSERVE" ANSWERS TO A PRINTED PROGRAM, AND THE RELATION OF ANSWER OBSERVING TO ERROR RATE. THIRTY COLLEGE STUDENTS STUDIED TWO PROGRAMS ON MEDICINE. FREQUENCY OF ANSWER OBSERVING WAS OBTAINED BY SELF-REPORT. PAY INCENTIVES (THE INDEPENDENT VARIABLE) WERE GIVEN TO ONE EXPERIMENTAL GROUP IMMEDIATELY BEFORE BEGINNING THE SECOND PROGRAM, AND TO ANOTHER IMMEDIATELY AFTER COMPLETION OF THE SECOND PROGRAM BUT BEFORE A POST TEST. IT WAS FOUND THAT THE PROBABILITY OF ANSWER OBSERVING ON ITEMS ANSWERED INCORRECTLY WAS HIGHER THAN ON FRAMES ANSWERED CORRECTLY BY AN INDIVIDUAL SUBJECT. INCENTIVES HAD AN INCONCLUSIVE EFFECT ON POST TEST SCORES. THIS PAPER APPEARED IN "STUDIES IN LANGUAGE AND LANGUAGE BEHAVIOR," PROGRESS REPORT V, SEPTEMBER 1, 1967. (LH)

Answer Observing in Programmed Instruction: II. The Effect of Incentive on Student Performance Within and After Programmed Instruction[1]

George L. Geis and Susan Nielsen

Center for Research on Language and Language Behavior
University of Michigan

30 college students served as paid Ss in a study of the frequency of answer observing while doing self-instructional programmed material. 2 different programs were used. Ss in the experimental groups were instructed either before doing the second program, or after, that they would be paid on the basis of their score on a post-program test. Frequency of error, frequency of answer observing, and post-test scores were determined. Ss were found to vary in their answer-looking frequencies; the probability of answer observing on those items (frames) which S had answered incorrectly was higher than on frames which he had answered correctly. The incentive (experimental) variable did not have a marked effect on any of the dependent variables except post-test score.

This is a continuation of a series of studies aimed at investigating those conditions in programmed teaching material which seem necessary if exposure to the correct answer of a frame is to be reinforcing to the student. In an earlier paper (Geis et al., 1965) the widespread contention that the printed answer in a text program is reinforcing to the student was questioned both logically and empirically. Those studies revealed that only under certain conditions did students observe answers in a program; a reasonable extension of that finding would be that the printed answer in a program is <u>reinforcing</u> to a student only when a specific set of conditions obtain. The pilot studies indicated that the population of learners was trimodal in regard to answer observing: a few students looked at every answer, a few students looked at no answers, and most students looked at about 1/4 to 1/3 of the answers that were available for observation. The studies also indicated that answer observing was related in a regular way to student error: the probability of a student observing an answer after having made an error on a frame was higher than after having made a correct response. Furthermore, one study revealed that the student's rating of his confidence in his own answer on a particular frame was as good a predictor as (or, perhaps, a better predictor than) the correctness of the student's answer.

The present study attempts to replicate the earlier findings and to investigate the effect on a number of dependent variables of two operations which might be termed incentive producing. The incentive instructions were given

to one group before they did a second program and to a second group immediately after they finished the second program but before they took the post-program test. Thus, these twogroups might be termed the "incentive learning" and the "incentive performance" groups.

Apparatus and Procedure

Materials. The materials used in the study were sections of two self-instructional programs designed as a review for physicians: the first 66 frames of *Allergy* and *Hypersensitivity* and the first 99 frames of *Physician's Liability* (Pfizer,[2] 1964, 1966).

Each frame was mounted separately on a 5" x 8" white index card; the answer to the frame was mounted on the reverse side of the card; the card was then enclosed in a clear plastic protector.

Some minor changes were made in the programs. In six frames of the *Liability* program a technical medical term referring to procedures or operations was accompanied by a more common synonym which was penciled in next to the technical word. One sentence which referred the reader to supplementary material in the appendix was removed from both programs. (These were additional references and, as noted in the text, were unnecessary for learning the material in the program.) Frame 30 in *Liability* could be considered calling for an opinion; no correct answer was given in the program. The frame was left in but neither errors nor looking data from that frame are considered in the results. In the few cases of continuous text material in the programs, materials which called for no written student response, the text was divided into convenient segments and mounted on successive cards in the same manner as the frames.

Accompanying each program was a blank answer sheet with item numbers and answer spaces duplicating those in the frame.

In addition, the apparatus included two cardboard boxes, one marked in large letters CHECKED and the other, UNCHECKED.

Three sets of frames and boxes were available for each program. Pre- and post-tests were constructed for the programs. The pre-test consisted of a 32-item vocabulary matching test in which single-sentence definitions were to be matched to technical terms. The definitions and terms were drawn from four areas: allergy, liability, linguistics, and thyroid function.

The *Allergy* post-test contained 20 items with a possible total score of 45 points; the *Liability* post-test had 24 items with a total score of 75 points. Each test consisted of three parts: constructed answer, fill-in, and multiple choice questions.

In addition each S was given a Need Achievement and Test Anxiety test, the data from which are not reported in this study.

Each S was given a short questionnaire to determine his academic background in relevant subject matter areas and experience with programmed instruction. Experimental sessions were held at the Center. Thirty Ss were used in the study. All were male freshmen or sophomores at the University of Michigan. Ss were paid in the manner described below.

Method

When a S reported for the first session, he was given a version of the Thematic Apperception Test (TAT) with accompanying written instructions. Upon completion of the TAT each S was given the Test Anxiety Questionnaire (TAQ), again with written instructions. Following this he filled out the short academic biographical form and took the pre-test. This concluded the first session. The S was paid for his time at the rate of $1.25/hr. and informed that he would be notified for a later session if he was to be used again.

In the second session, the S was seated before a table on which was a set of program cards and the pair of labelled boxes. He was then read the instructions:

This is an experiment in programmed instruction. Each index card presents a frame which provides information and may require you to make a response. You will be supplied with sheets on which to write your responses.

The answers to each frame are on the reverse side of the card. If you wish to check your response against the correct answer, simply flip the card over after you have written your answer on the answer sheet. You may check as many or as few responses as you wish.

If you check an answer by flipping over the card, place the card in the box marked "checked." If you do not check the answer, place the card in the box marked "unchecked." If you decide to check the printed answer

to the frame and you find that you are incorrect, do not go back and change your response. Do not go back to previous frames unless you are specifically instructed to do so. You will be given a post-test when you have finished.

Are there any questions?

Half of the Ss went through the *Liability* program and half through the *Allergy*. Ss worked at separate tables usually in small groups of three or four. The time taken to complete the programs was recorded.

Upon the completion of the program the S was given the post-test, one page, or section, at a time. Before leaving the session a time was set for the next session. The S was not paid at the end of this session, having been informed that he would receive full pay at the end of his last session.

At the beginning of Session 3, Ss were instructed differently according to their grouping. Three groups had been constructed, balancing each for distribution of TAT and TAQ scores. Each S in Session 3 worked on the program he had not done in Session 2.

Groups I, II, and III were re-read the instructions for Session 2.

Group II Ss ("incentive learning" group) were then given typed instructions which had been clipped to the first page of their program response sheets:

You have already earned $____ in the previous session. Your final payment for this session will depend upon how well you score on the post-test. You will receive 5¢ for each percentage point of your score. Therefore, you can earn as much as $5.00 for this session. I cannot tell you any more than this. After completing the program, Group III Ss ("incentive performance" group) were given typed instructions which had been clipped to the first page post-test answer sheet:

Now that you have finished the program, here are some further instructions.

You have already earned $____ in the previous session. Your final payment for this session will depend upon how well you score on the post-test. You will receive 5¢ for each percentage point of your score. Therefore, you can earn as much as $5.00 for this session. I cannot tell you any more than this.

Each S went through the program, took the post-test and then was paid for his last two sessions. Before leaving he was given instructions not to discuss the experiment with other students and was thanked for his cooperation.

A summary of the groups and procedures may be found in Table 1.

Insert Table 1 about here

Results and Discussion

Previous research has led to the expectation that with most programs a small number of the frames and Ss contributes most of the errors made on frames. Figures 1 and 2 indicate this to be true of the two programs and the populations used in the present study. In Figure 1 the cumulative error is plotted against the frames in the programs ranked for error. Thus, the first point plotted

---

Insert Figure 1 about here

---

for the *Liability* program is the total number of errors recorded for the highest error frame, the second point adds to it the total number of errors scored on the second highest error frame, etc. More than half of the total number of errors made by all Ss is accounted for by less than 1/4 of the frames in each program. In Figure 2 it can be seen that about 1/3 of the Ss account for half of the errors recorded.

Insert Figure 2 about here

The per cent error for Ss and frames is higher than that reported as generally acceptable in well-designed programs. However, these programs were designed for physicians and our test population consisted of college students with no medical training. The material was not only foreign to them (as revealed by low scores on pre-tests), but also the interest or motivation for doing the program in this population was probably based upon payment whereas in the target population there are professional reasons for the physician to master the material in the programs. Therefore, it would be a mistake to be critical of the programs *qua* programs on the basis of the high per cent error indicated here. Nevertheless, the pattern if not the magnitude, of high error frames and high error Ss is typical of the data we have obtained on other programs.

Figure 3 shows frequency plots of answer observing for each program, the data being presented separately for first and second program presentations.

Insert Figure 3 about here

Thus, *Liability* Ia, IIa, IIIa is a figure of the data of those 15 Ss who took the *Liability* program first and the *Allergy* program second. In all cases the data are similar to those obtained in previous studies (see Figure 4). Students

Insert Figure 4 about here

do not regularly observe all of the answers in a program. On the average, they looked about one-half of the time on these programs and, in the case of all but one group, regularly observed answers more often on the second program than on the first. As in previous studies, there are Ss who never look at an answer in the program, Ss who always look at an answer in the program, and the majority are somewhere in the middle.

As previously mentioned, earlier studies revealed the relationship between student error and answer observing. Figures 5 and 6 present data relevant to this observation. In these figures the probability of observing an answer after

Insert Figures 5 and 6 about here

having made an incorrect response (Pw/w) is plotted against the total percentage of looks for each S. The diagonal line on each figure represents chance level of looking when wrong for all total answer-looking per cents. Thus, if a S looked at answers in the program 50% of the time, his chance level of looking when wrong would be 50% and his data point would fall on the line. It is obvious that in almost all cases the Ss, regardless of their total per cent of looking, look at answers in the program much more often when they are incorrect than when they are correct; almost all of the data points fall to the left of the indicated chance level. The implication is, of course, that the S is able to discriminate, on the basis of his own answer, something about the correctness of that answer and tends to check the printed answer in those cases in which he himself has supplied an incorrect answer. A second observation may be made about these data--namely, that looking and looking-when-wrong percentages are related. Figures 5 and 6 are not good illustrations of this point since the total looking rate is plotted on the abscissa. However, when the data are analyzed into looking-when-right and looking-when-wrong rates, a clear relationship between the rates is disclosed, i.e., extremely high Pw/w

predicts high looking rates when correct and vice versa while low Pw/w predicts low Pc/c and vice versa. The frequency of looking-when-wrong is not a function of the error rate of a particular S. A low error-rate S may or may not have as good a discrimination (a tendency to look more often when wrong than when right) as a high error S.

It is interesting to note from data not graphically presented that the Pw/w is generally greater for the second program than for the first. There is at least an indication here that as a student continues to work on the programs he tends to develop more effective or efficient strategies. The most efficient strategy, of course, would be to check only those answers which are incorrect. This indication of the acquisition of more effective learning strategies as a function of more experience with the learning materials will be pursued in later studies. The word "effective" is intentional, as there is some indication from the data that a shift upward in a student's Pw/w is accompanied by an increased post-test score.

Figure 7 shows consistency of looking within Ss. A significant correlation is obtained when the per cent looking on a first program is plotted as a function of the per cent looking on the second program. Thus, a S can truly be termed a high or low looker.

Insert Figure 7 about here

Claims have been made for the necessity of observing the answers in a program in order to produce effective learning. Figure 8 shows the relationship or lack of it between answer observing and post-program test score

---

Insert Figure 8 about here

---

using the *Allergy* program. No relationship is found. A simple-minded notion that mere exposure to the answer as well as to the frame on a program is important in learning must therefore be discarded.

It has already been pointed out that the error rates are high for some Ss and for some frames in each program. The consistency of error within Ss is demonstrated in Figure 9. Here the per cent errors in the two programs are

---

Insert Figure 9 about here

---

plotted. It seems safe to conclude that there are Ss who consistently maintain a high or low error level regardless of the program. The picture emerges from these observing and error data that the within-S variables such as strategy of learning and competencies related to performance within the program show up regularly and consistently. These variables clearly deserve further study.

Figure 10 lends some support t the conjecture that errors on programs are related to the amount of achievement in the program. Here the post-program test scores for the *Allergy* program are plotted against the per cent error within the program. The lower error Ss clearly tend to perform better on the test.

Insert Figure 10 about here

Group comparisons. Until now the data have been reported with all Ss considered as a single group. Table 1 presents the means for Ss in different groups for the various dependent variables. Groups Ia and Ib might be considered control groups. The *Liability* post-test scores tend to be higher than the *Allergy* post-test scores. Data from the experimental groups indicate that the instructions regarding post-test contingent payment, whether those instructions were given before or after the Ss went through the program, tended to raise the post-test scores. Thus, the *Allergy* post-test score seems inflated for Group IIa; the *Liability* post-test score seems inflated for Group IIb and similarly, the increase in the test score related to the payment instructions can be found in the two sub-groups of Group III. If the difference scores are calculated for the *Liability* minus the *Allergy* post-tests for each group, it is clear that instructions seem to have tipped the balance in favor of the payment-connected post-test. It is not unexpected that the setting of such explicit contingencies would produce a change in the achievement scores. The difference between being instructed about the contingencies before taking the program or directly before taking the post-test (the difference between Groups II and III) seems to be insignificant. The difference scores for Group IIb and IIIb are exactly the same, the scores for Groups IIa and IIIa are only slightly different.

Of greater interest in this study was the effect of instructions on answer observing. It was predicted that Group II Ss would tend to observe more answers than Groups I or III. The simple but logical hypothesis was that

the instructed Ss in Group II might tend to be more cautious and develop a strategy which involved a higher rate of answer observing. Again, difference scores in observing suggest that Group II did observe relatively more on the program following instructions than would be expected, extrapolating from the Group I data. However, the data from Group IIIb casts a shadow upon any conclusion concerning the effectiveness of instructions.

Data from the control group suggest that answer observing is a function of interaction between the program content and whether the program is taken first or second. The control group shows that the *Liability* program tends to produce more answer observing. Thus, the difference scores between answer observing for *Liability* and *Allergy* for Group Ia is -.6 and for Group Ib, +10.0. The tentative conclusion is that the *Liability* program produces relatively more observing behavior than the *Allergy* program when the factor of order is removed. The prediction might then be that for the sub-groups in Group II these tendencies would be exaggerated. Thus, in Group IIa the *Allergy* program was in second position as well as the program before which instructions were given. Therefore, the difference between observing on *Allergy* and *Liability* should be increased in favor of *Allergy*. This is indeed what occurred. In Group IIb the *Liability* program was second and also was the pre-instructed program; the difference should be exaggerated in favor of the *Liability* program. In this case, the difference score is not as great as in Group Ib. In the sub-groups of Group III there is no reason to expect data that differ markedly from those of Group I and yet they do. Group IIIb especially shows an unexpected reversal with the second program (*Liability*) having less answer observing than the first program (*Allergy*). There is some reason to suspect that Group III was an atypical collection of Ss. But rather than attempt to argue away the data, it should merely be noted that there is little support for any conclusion which suggests the instruction variables had a marked effect on answer observation.

With regard to errors on the program, the error rate on the *Liability* program is consistently higher than the error rate on the *Allergy* program for all groups. When the difference scores are determined, by subtracting the error rate for *Allergy* from the error rate for *Liability*, again a

comparison can be made between instructed and non-instructed groups. It might be expected that Ss in Groups IIa and IIb would show lower error scores because of the care with which they would go through the program, knowing ahead of time that their post-test scores would determine their payment. Thus, in Group IIa the expected smaller *Allergy* error rate should be even further reduced while in Group IIb the higher *Liability* error rate should be reduced. Again, the trend is in the right direction with an increase in the difference score from Group IIa and a decrease in the difference score for Group IIb, but the variability of the data precludes any conclusion about the effect of the instructional variable on the error rates within the program.

Summary and Conclusion

The study has replicated and supported earlier indicative data concerning the variability in rates of observation of answers in a program, and the relationship between answer observation and error in programs. In addition, it has provided data which indicate there is within-S consistency in terms of error and in terms of answer observing. The present study supports the observation previously made on the basis of earlier studies that the answer in a program cannot automatically be considered reinforcing but rather may act as a reinforcer only under certain conditions.

Reference

Geis, G. L., Jacobs, W., & Spenser, D. The role of the printed answer in programmed instruction. *Studies in language and language behavior, I*, October 15, 1965, Center for Research on Language and Language Behavior, University of Michigan, Contract No. OE5-14-036, U. S. Office of Education.

Footnotes

[1]The research reported herein was performed pursuant to Contract OEC-3-6-061784-0508 with the U. S. Department of Health, Education, and Welfare, Office of Education, under the provisions of P. L. 83-531, Cooperative Research, and the provisions of Title VI, P. L. 85-864, as amended. This research report is one of several which have been submitted to the Office of Education as *Studies in language and language behavior*, Progress Report V, September 1, 1967.

[2]The authors wish to thank the Charles Pfizer & Co., Inc. for providing copies of these programs.

Table 1

Mean Post-Test Answer-Looking and Frame Errors and Ranges for Sub-Groups

GROUP	PROGRAM (in order of presentation)	PAYMENT INSTRUCTION VARIABLE	POST-TEST MEAN	RANGE	ANSWER LOOK MEAN (RANGE)	ERROR MEAN (RANGE)
Ia	LIABILITY	no post-test payment instruction	87.6	95-76	54.8 (78-42)	30.6 (49-23)
	ALLERGY	no post-test payment instruction	81.2	91-67	55.4 (100-26)	25.2 (41-15)
Ib	ALLERGY	no post-test payment instruction	83.8	93-71	45.6 (95-5)	19.2 (27-15)
	LIABILITY	no post-test payment instruction	93.2	99-87	55.6 (100-5)	26.6 (35-23)
IIa	LIABILITY	no post-test payment instruction	91.4	97-83	57.4 (82-22)	25.2 (37-6)
	*ALLERGY	instructions on payment before program	87.8	98-78	71.2 (92-40)	18.2 (30-8)
IIb	ALLERGY	No post-test payment instruction	77.2	93-56	78.8 (100-23)	20.6 (33-14)
	*LIABILITY	instructions on payment before program	93.4	98-89	87.4 (100-37)	21.2 (32-3)
IIIa	LIABILITY	no post-test payment instruction	90.4	99-70	47.2 (89-0)	22.4 (38-12)
	*ALLERGY	instructions on payment before test	94.0	98-82	53.8 (95-0)	15.2 (27-9)
IIIb	ALLERGY	no post-test payment instruction	77.8	96-47	58.6 (100-14)	19.8 (35-11)
	*LIABILITY	instructions on payment before test	94.0	96-89	44.0 (80-11)	21.6 (32-4)

*Experimental groups.

Figure Captions

Fig. 1. For each program the cumulative number of errors on program frames is plotted as a function of frames ranked for errors. (Data are for 30 Ss.)

Fig. 2. For each program the cumulative number of errors on program frames is plotted as a function of Ss ranked for error.

Fig. 3. Frequency plots are shown for the number of Ss observing answers at each observing-frequency interval. Each group consisted of five Ss; Groups Ia, IIa, IIIa took the *Liability* program first; Groups Ib, IIb, IIIb took the *Allergy* program first.

Fig. 4. Frequency plots of answer-observing are shown using data from four previous studies.

Fig. 5. Each S's per cent of answer observing when wrong (Pw/w) is plotted as a function of his total per cent answer observing in the *Allergy* program.

Fig. 6. Each S's per cent of answer observing when wrong (Pw/w) is plotted as a function of his total per cent of answer observing in the *Liability* program.

Fig. 7. Each S's total per cent of answer observing in the *Liability* program is plotted as a function of his total per cent of answer observing in the *Allergy* program.

Fig. 8. For the *Allergy* program, each S's total per cent of answer observing is plotted as a function of his score (per cent) on the *Allergy* post-test.

Fig. 9. Each S's per cent of frame errors in the *Liability* program is plotted as a function of his per cent frame errors in the *Allergy* program.

Fig. 10. Each S's post-test score (%) is plotted as a function of his frame errors in the *Allergy* program.

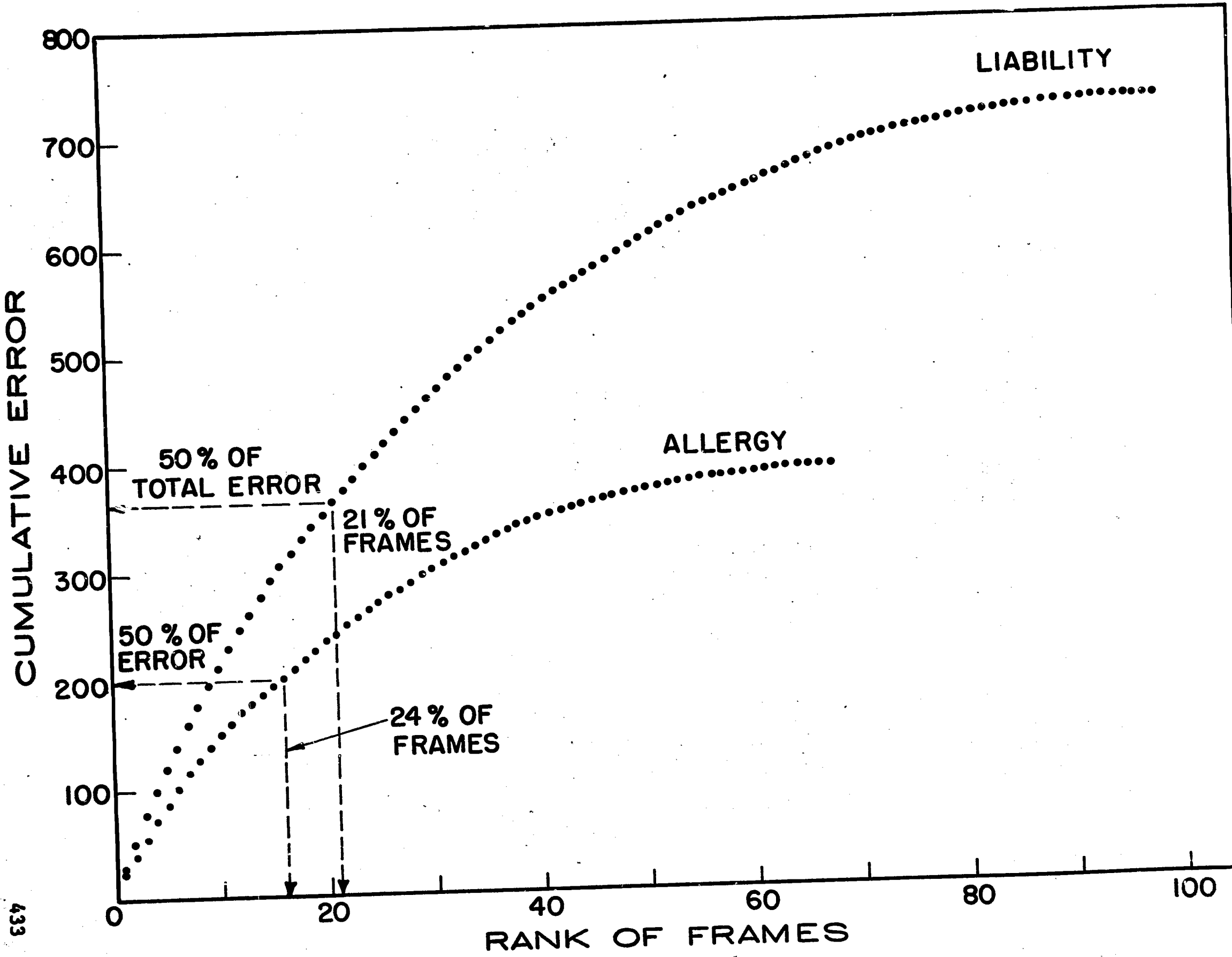
LIABILITY
ALLERGY
CUMULATIVE ERROR
800
700
600
500
400
300
200
100
0
20
40
60
80
100
RANK OF FRAMES
50 % OF
TOTAL ERROR
21 % OF
FRAMES
50 % OF
ERROR
24 % OF
FRAMES

Figure 1

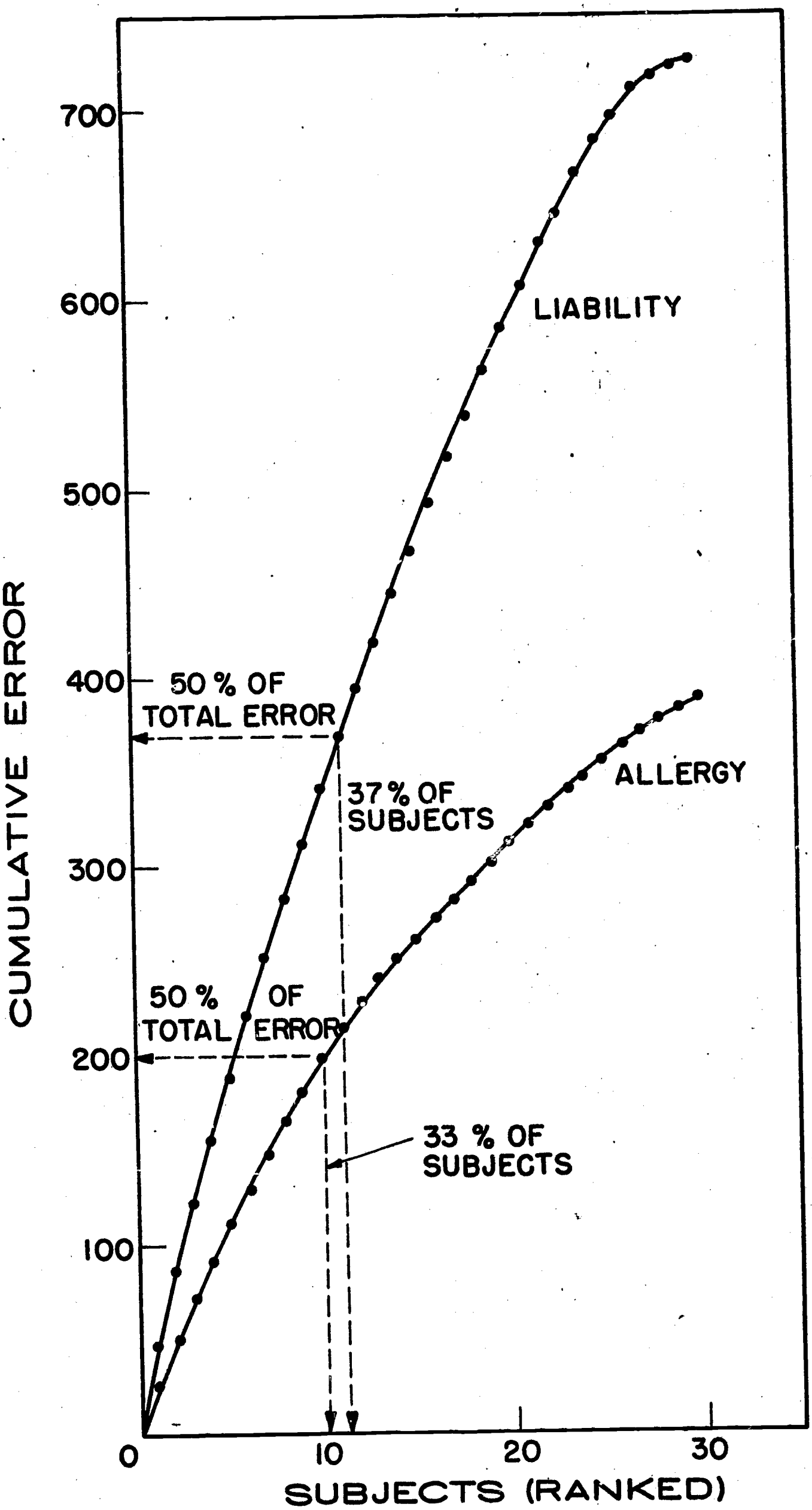
700
600
500
400
300
200
100
0
10
20
30
CUMULATIVE ERROR
SUBJECTS (RANKED)
LIABILITY
ALLERGY
50 % OF TOTAL ERROR
37 % OF SUBJECTS
50 % OF TOTAL ERROR
33 % OF SUBJECTS

Figure 2

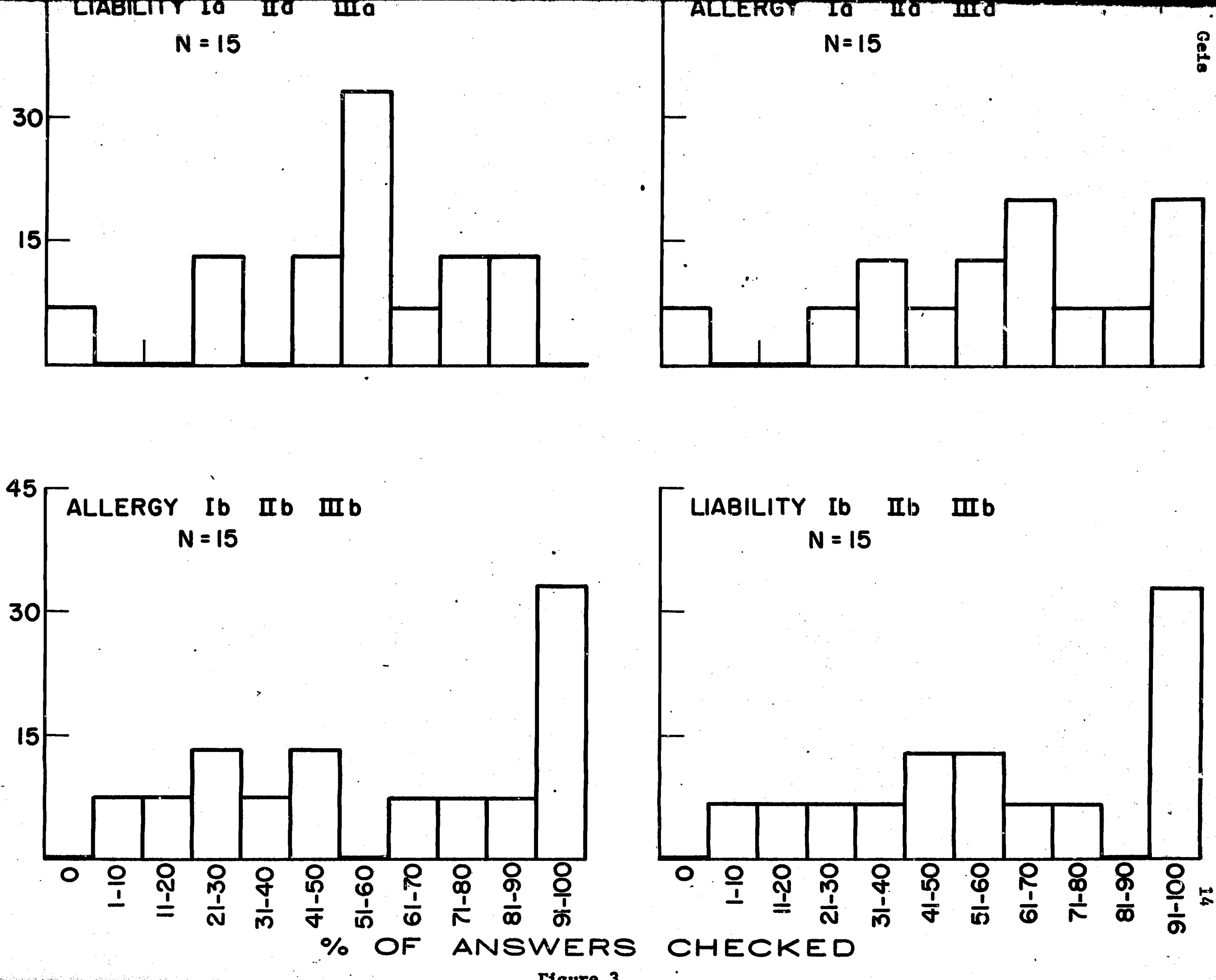

Figure 3

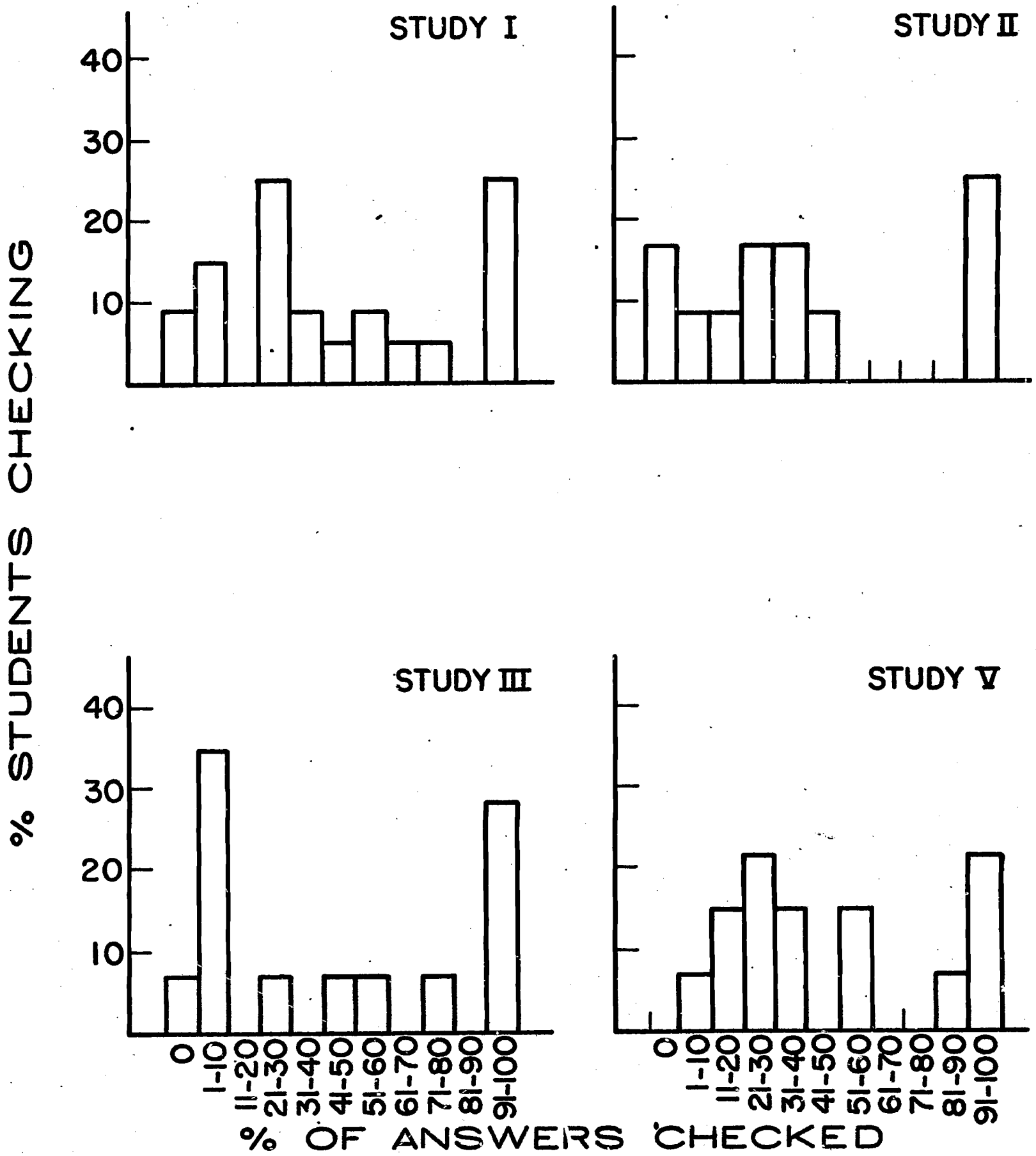
STUDY I
STUDY II
STUDY III
STUDY V
40
30
20
10
% STUDENTS CHECKING
0
1-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100
% OF ANSWERS CHECKED

Figure 4

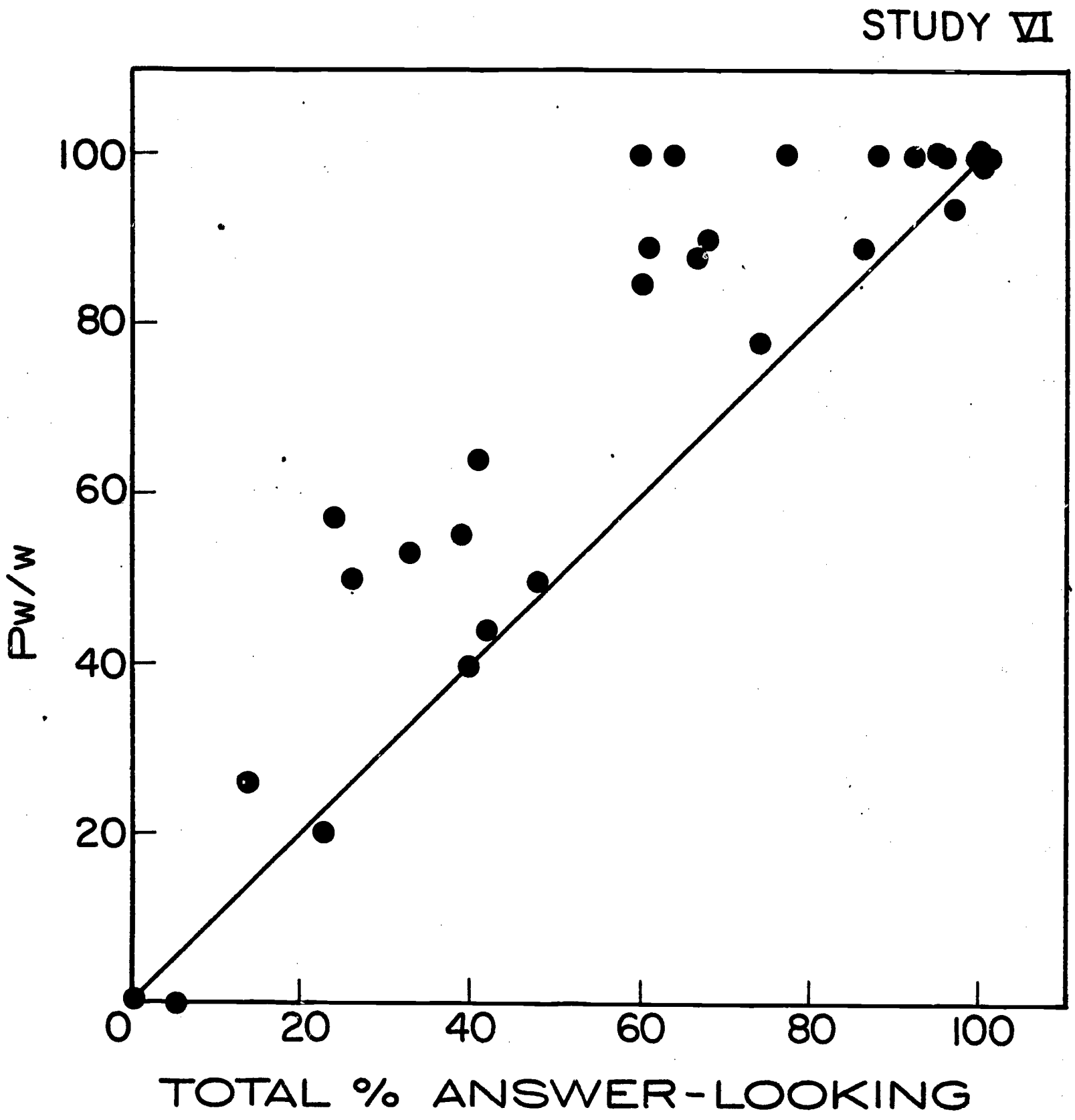

Figure 5. Allergy

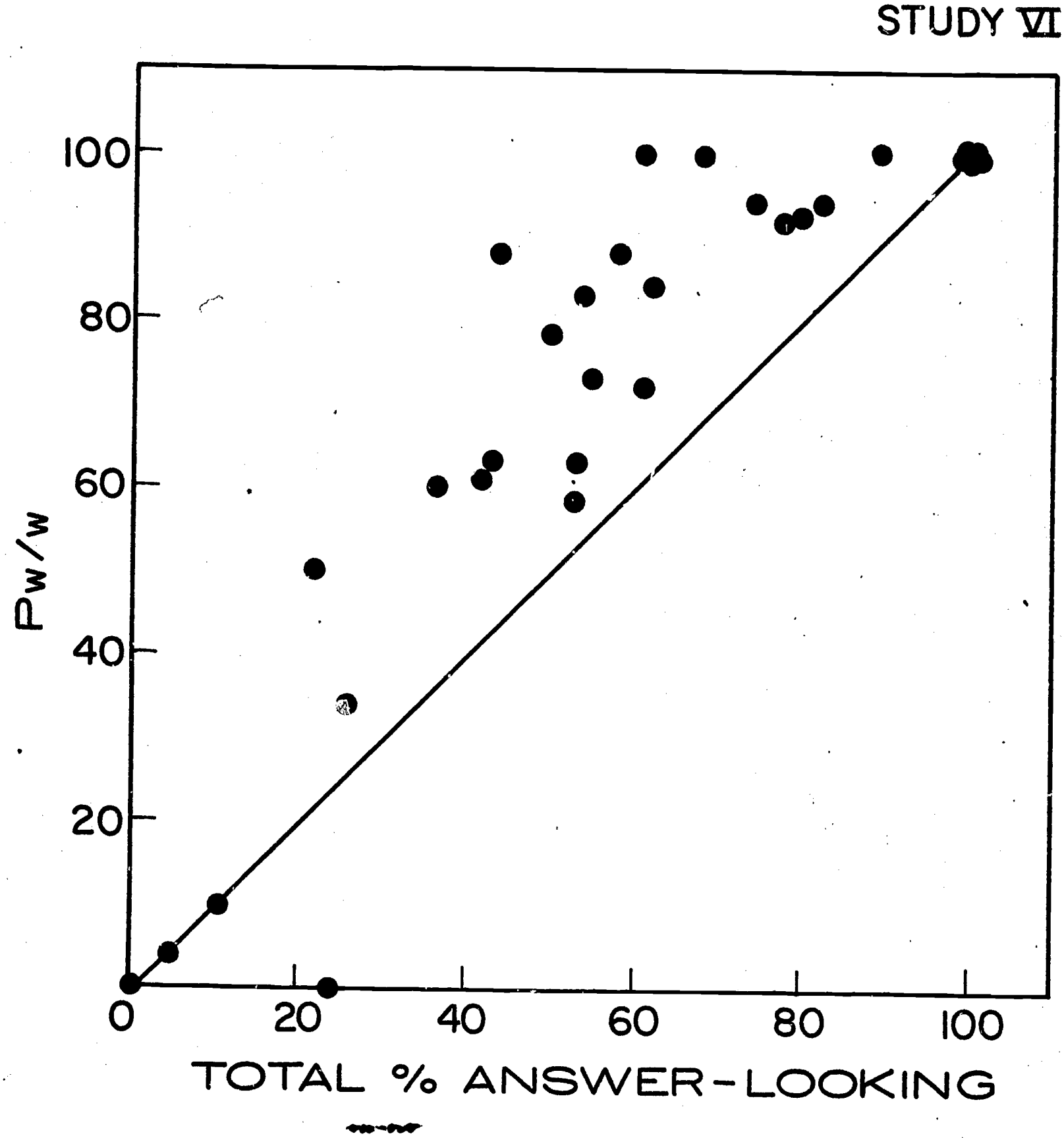

Figure 6. Liability

STUDY VI

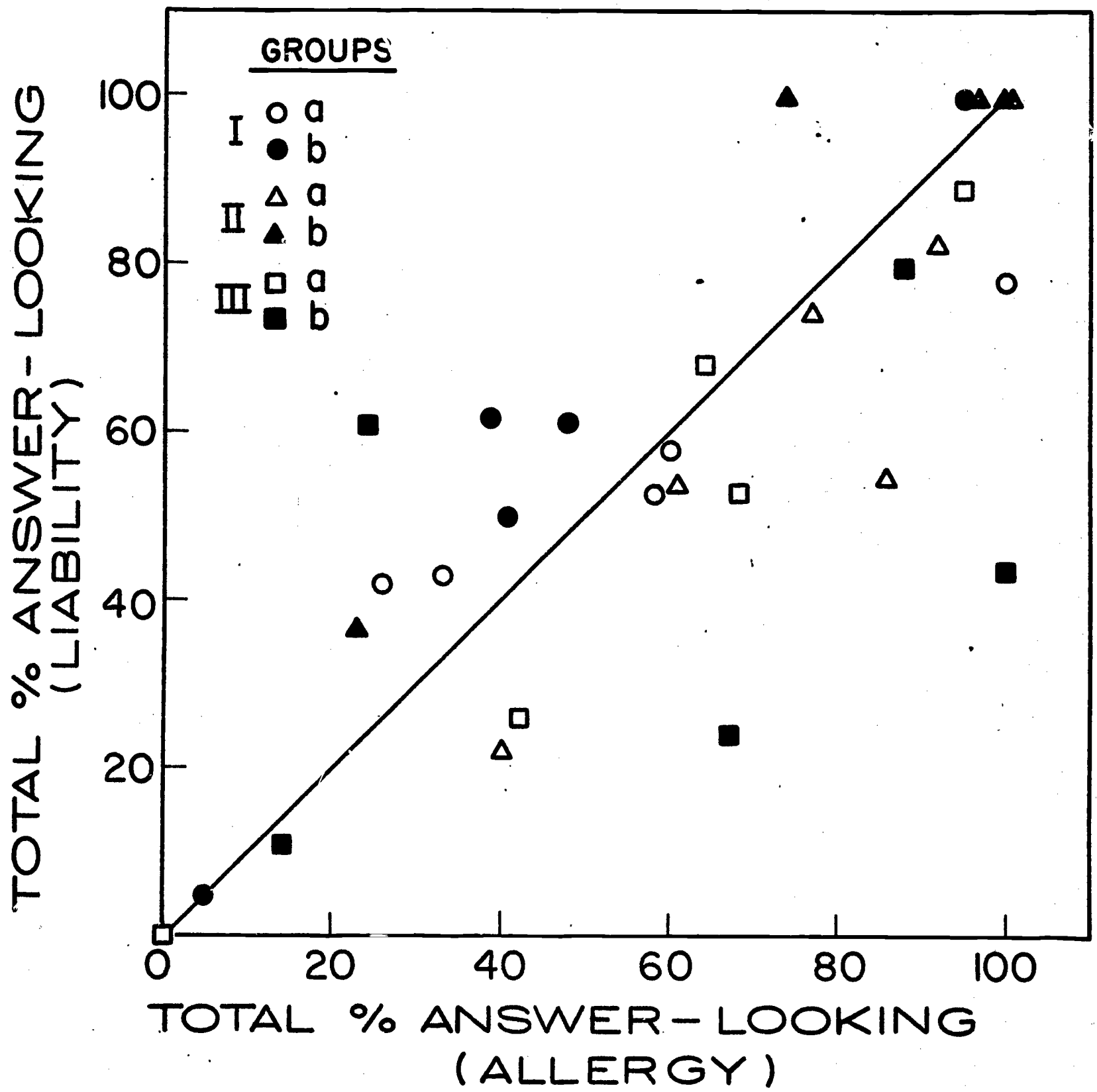

Figure 7

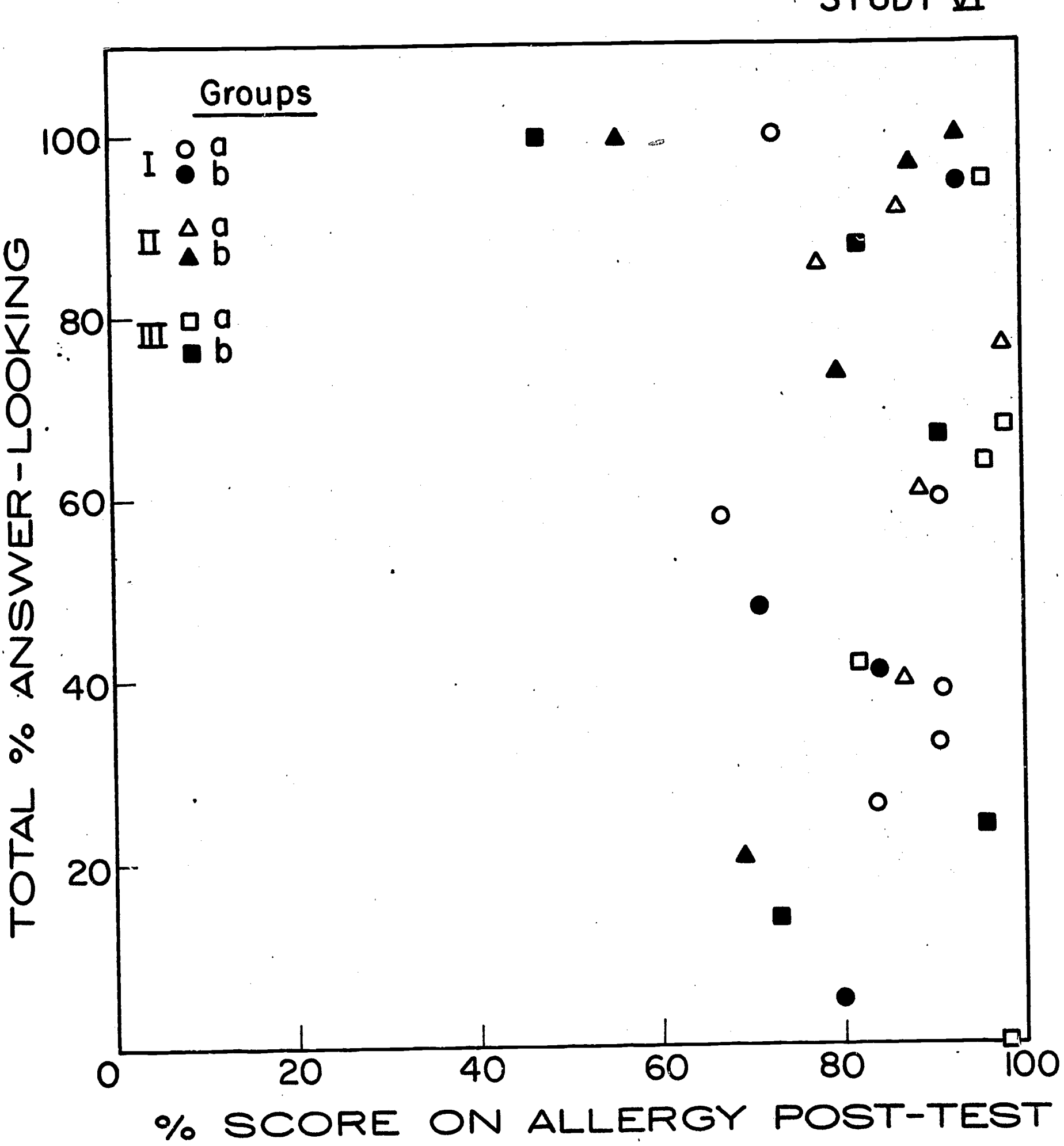
STUDY VI
Groups
I a
b
II a
b
III a
b
100
80
60
40
20
0
20
40
60
80
100
TOTAL % ANSWER-LOOKING
% SCORE ON ALLERGY POST-TEST

Figure 8

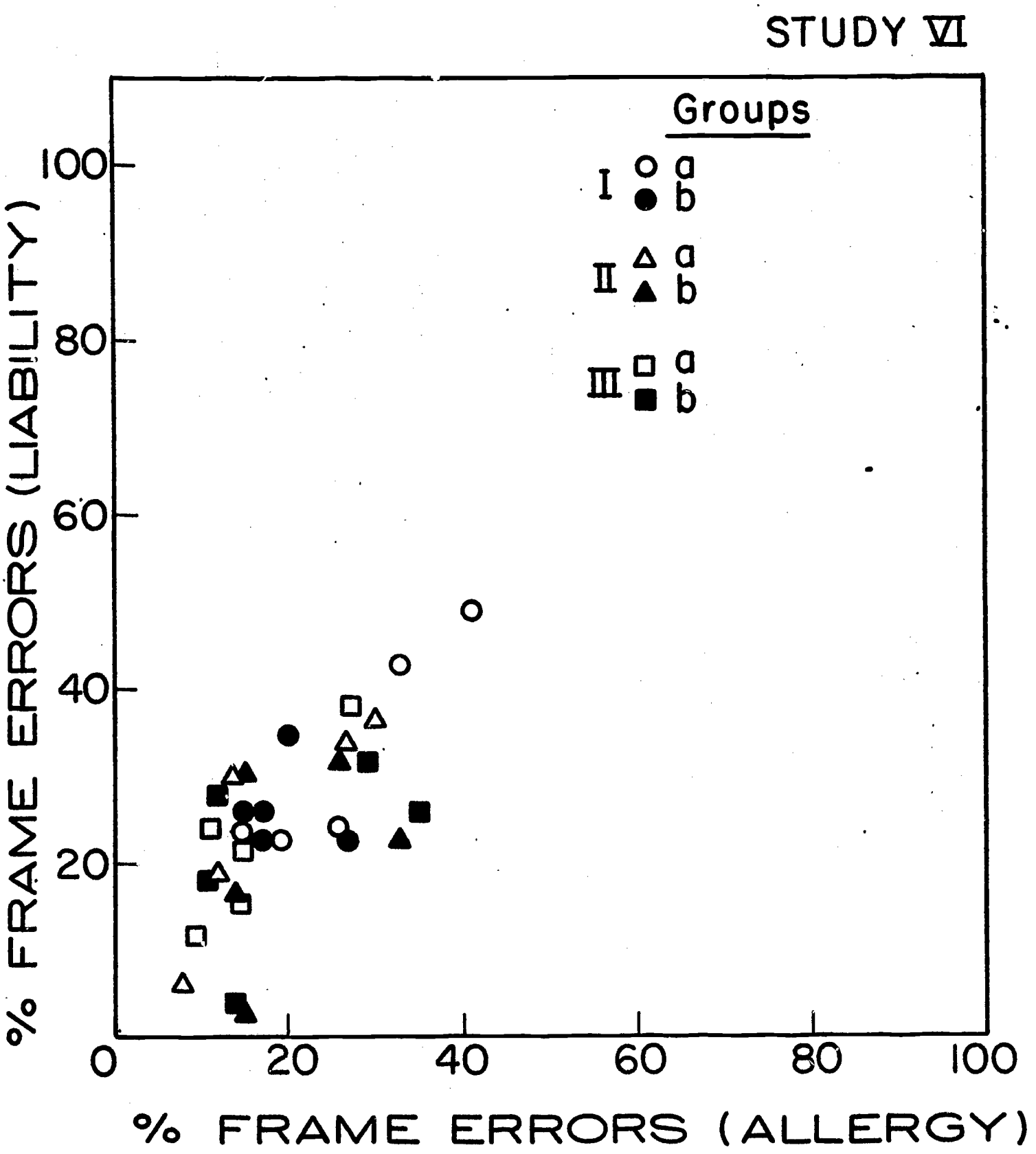
STUDY VI
Groups
I a
b
II a
b
III a
b
100
80
60
40
20
0
20
40
60
80
100
% FRAME ERRORS (LIABILITY)
% FRAME ERRORS (ALLERGY)

Figure 9

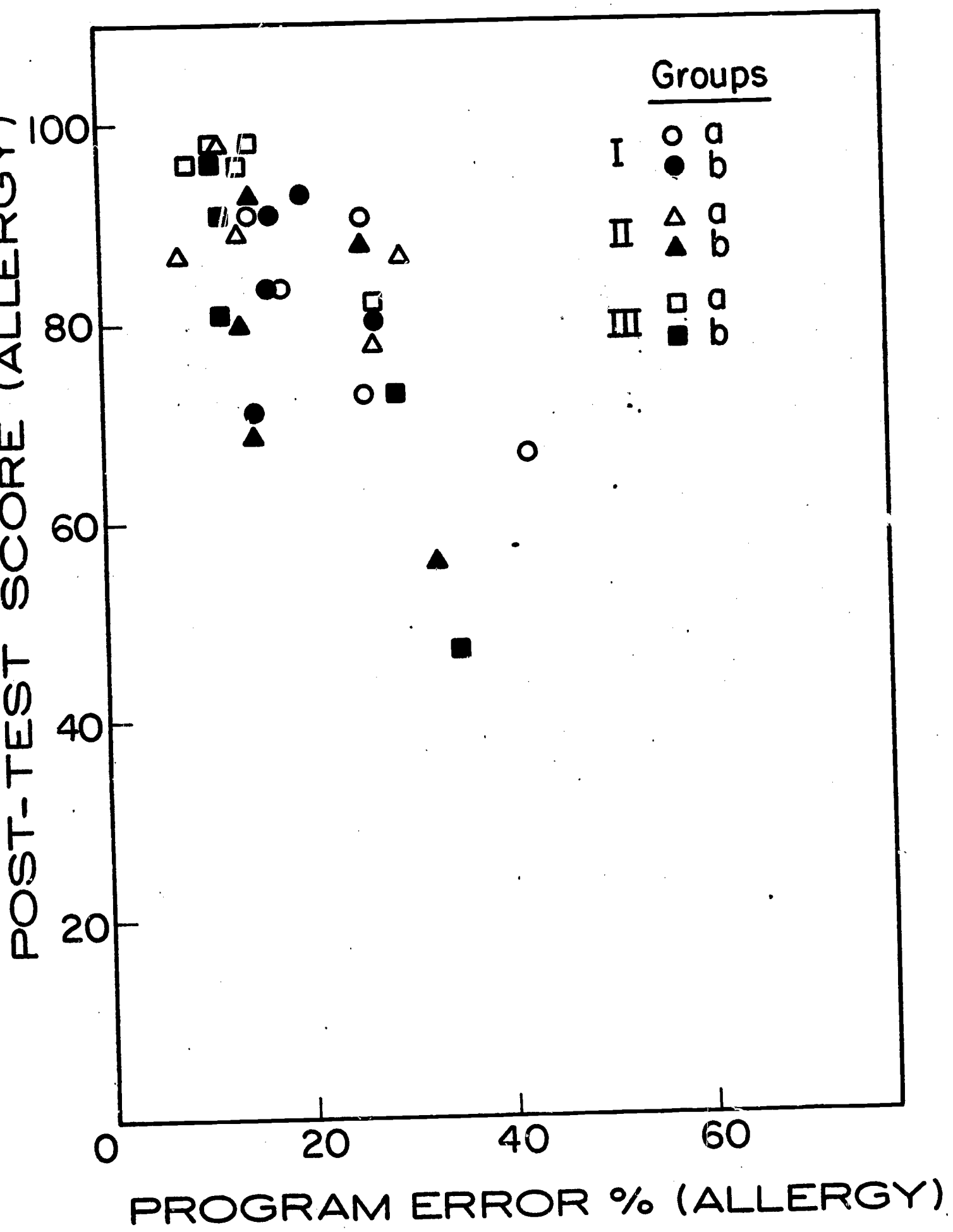
STUDY VI
Groups
I a
b
II a
b
III a
b
POST-TEST SCORE (ALLERGY)
100
80
60
40
20
0
20
40
60
PROGRAM ERROR % (ALLERGY)

Figure 10